2015

Stochastic Performance Trade-offs in the Design of Real-Time Wireless Sensor Networks

Yunbo Wang
University of Nebraska-Lincoln, yunbowang@gmail.com

Mehmet C. Vuran
University of Nebraska-Lincoln, mcvuran@cse.unl.edu

Steve Goddard
University of Nebraska – Lincoln, goddard@cse.unl.edu

2015 International Conference on Computing, Networking and Communications, Wireless Ad Hoc and Sensor Networks Symposium Pages: 931 - 937, DOI: 10.1109/ICCNC.2015.7069471

Stochastic Performance Trade-offs in the Design of Real-Time Wireless Sensor Networks

Yunbo Wang Mehmet C. Vuran Steve Goddard
Department of Computer Science and Engineering
University of Nebraska-Lincoln
Email: yunbowang@gmail.com, {mcvuran, goddard}@cse.unl.edu

***Abstract*—Future sensing applications call for a thorough evaluation of network performance trade-offs so that desired guarantees can be provided for the realization of real-time wireless sensor networks (WSNs). Recent studies provide insight into the performance metrics in terms of first-order statistics, e.g., the expected delay. However, WSNs are characterized by the stochastic nature of the wireless channel and the queuing processes, which result in non-deterministic delay, throughput, and network lifetime. For the design of WSNs with predictable performance, probabilistic analysis of these performance metrics and their intrinsic trade-offs is essential. Moreover, providing stochastic guarantees is crucial since each deployment may result in a different realization.**

In this paper, the trade-offs between delay, throughput, and lifetime are quantified through a stochastic network design approach. To this end, two novel probabilistic network design measures, *quantile* and *quantile interval*, are defined to capture the dependability and predictability of the performance metrics, respectively. Extensive evaluations are conducted to explore the performance trade-offs in real-time WSNs.

I. Introduction

Wireless sensor networks (WSNs) have been utilized in many applications as both a connectivity infrastructure and a distributed data generation network due to their ubiquitous and flexible nature. Increasingly, a large number of WSN applications are investigated with various real-time performance requirements for different network services specific to low-cost hardware and unpredictable environmental conditions [1]. These requirements necessitate a comprehensive analysis of the real-time performance guarantees provided by the network.

In this paper, we explore and quantify the probabilistic performance trade-offs in the design of real-time WSNs with an anycast protocol. More specifically, we consider the trade-offs between end-to-end communication delay, the network lifetime, and throughput of the network. To quantify the *dependability* of a probabilistic real-time network, a *quantile*-based measure is defined, which defines the end-to-end delay or the network lifetime that can be achieved with at least probability p. Moreover, to quantify the *predictability* of a network, a *quantile interval*-based measure is defined, which captures the difference in end-to-end delay or network lifetime between two quantiles p_1 and p_2. We aim to answer questions such as: how does the maximum network lifetime change if we want to improve the predictability of the end-to-end delay? If we require such a network to operate for longer than 6 months with at least a probability of 0.8, what would be the minimum network density to satisfy these requirements? What are the tradeoffs between stochastic requirements of network lifetime and end-to-end delay? To the best of our knowledge, this work is the first to quantify the probabilistic performance trade-offs in WSNs.

The remainder of this paper is organized as follows: In Sec. II, the models used to derive the stochastic end-to-end delay and network lifetime for an anycast protocol is briefly described, and related work is discussed. Then, in Sec. III, the problem definition is provided and the evaluation methodology is described. The evaluation results of the probabilistic analysis models and our major findings are presented in Sec. IV. Finally, the paper is concluded in Sec. V.

Supported, in part, by grants from the National Science Foundation (CNS-1117664, CNS-1247941, CNS-0953900, and CNS-1331895).

II. Background and Related Work

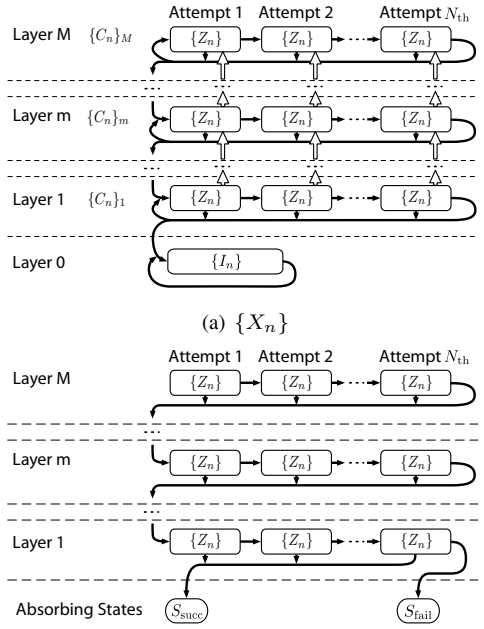
Compared to first-order performance statistics, the probabilistic distribution of a performance metric provides tools that can be leveraged to design networks with desired performance guarantees. In the following, we revisit the probabilistic distribution models that will be utilized in the remainder of the paper, and then related work in this area is discussed.

A. Probabilistic Distribution Models

Consider a network where nodes are deployed randomly in a 2-D circular plane of radius R, according to a Poisson point process. Each node senses the physical events, generates traffic with rate λ_{lc}, and then forwards the generated packets to a sink, located at the center of the plane, through multi-hop communications. Assume each node has a battery capacity, C. In our previous work [17, 20], models are developed to analyze the probability distribution of the end-to-end communication delay and the network lifetime in such networks. These models utilize a Discrete-Time Markov queueing model in node-level analysis and fluid models in network-level analysis.

Consider that each node is identified according to its location $\boldsymbol{x}$, and its performance is affected by three design parameters: the traffic generation rate, λ_{lc}, the network density, ρ, and the duty cycle, ξ. The anycast communication technique, which has been adopted in terrestrial, airborne, and underwater WSNs [2, 6, 12, 15], is considered for its efficiency in both delay and energy consumption. Moreover, a log-normal fading channel model is employed [23]. Due to space limitations, this section provides only a brief summary and overview of the models developed in [17, 20].

978-1-4799-6959-3/15/$31.00 ©2015 IEEE



(a) $\{X_n\}$

(b) $\{Y_n\}$

Fig. 1: (a) The layered structure of Markov chain for $\{X_n\}$ and (b) its absorbing version $\{Y_n\}$.

1) Discrete-Time Markov Process based Node-level Models: The node-level performance is modeled based on a Discrete-Time Markov Process (DTMP), in which time is divided into units of length T_u. Accordingly, the probability distributions of the single-hop delay between a pair of nodes $\boldsymbol{x}$ and $\boldsymbol{x}'$, the single-node energy consumption for node $\boldsymbol{x}$ during t, and the lifetime for node $\boldsymbol{x}$ are obtained.

In the DTMP-based analysis, each node is modeled according to a first-come-first-serve queuing system, which is characterized by a Quasi-Birth-Death (QBD) process [10], as explained in [20]. The discrete-time Markov chain (DTMC) used to represent the QBD process, denoted as $\{X_n\}$, has a layered structure, as shown in Fig. 1(a). Each layer $m \in \{0, \ldots, M\}$ contains the part of the chain where there are m packets in the queue and each state represents the activity that is conducted by the node during each time unit of T_{u}. First, the transition probability matrix, $\boldsymbol{Q}_X$, of the entire Markov chain $\{X_n\}$ is found. Then, the equilibrium state probability vector, $\boldsymbol{\pi}$, for $\{X_n\}$ is calculated by solving $\boldsymbol{\pi}\boldsymbol{Q}_X = \boldsymbol{\pi}$. The detailed Markov chain construction and solution are described in [20].

The *pdf* and the *cdf* of the energy consumption during T (in integer multiples of T_u) are [17, 19]

$$f_{E_{\text{cp}}(T)}(e) = \boldsymbol{\pi}\boldsymbol{h}^{(\hat{T})}(e)\mathbf{1},\ F_{E_{\text{cp}}(T)}(e) = \int_0^e f_{E_{\text{cp}}(T)}(\epsilon)\text{d}\epsilon, \quad (1)$$

respectively, where $\boldsymbol{h}^{(\hat{T})}(e)$ is given by (15) in [19], and $\mathbf{1}$ is the appropriately dimensioned column vector containing all 1's. It is also shown that when T is large, the total energy consumption for communication and data processing during T asymptotically approaches the Normal distribution [17]. Moreover, the energy consumption during a given time period, T, is expressed as the sum of three independent random variables: energy consumption for sensing, communication and processing, and an empirically determined zero-mean random variable that captures the randomness in energy consumption due to topology. Accordingly, the *pdf* of the total energy consumption of a node at $\boldsymbol{x}$ is obtained by (2) in [19], and the mean and variance of the asymptotic Normal distribution are given by (22) and (23) in [19].

To derive the single-hop delay distribution, another DTMC, $\{Y_n\}$, which is an absorbing variant of $\{X_n\}$, is used. The *pmf* of the single-hop delay for successful and failed communication, measured in number of time units, t_{sh}, are given as

$$f^{\text{s}}_{t_{\text{sh}}}(k) = \boldsymbol{\alpha}_Y \boldsymbol{P}_Y^{k-1}\boldsymbol{t}^{\text{s}}_Y,\ \ f^{\text{f}}_{t_{\text{sh}}}(k) = \boldsymbol{\alpha}_Y \boldsymbol{P}_Y^{k-1}\boldsymbol{t}^{\text{f}}_Y, \quad (2)$$

respectively, where $\boldsymbol{\alpha}_Y$, $\boldsymbol{t}^{\text{s}}_Y$, $\boldsymbol{t}^{\text{s}}_Y$, and $\boldsymbol{\alpha}_Y$ are obtained according to (14)-(16) in [20].

2) End-to-End Delay and Network Lifetime Distributions: The network lifetime is defined as the duration before the battery depletion of the first node. Based on the single-node energy consumption analysis, the network lifetime is obtained according to [17, 19].

With each hop modeled as a Geom/PH/1/M queue, the entire network is considered as a queueing network. Based on the single hop delay distribution for each pair of nodes, the end-to-end delay is obtained using an iterative procedure [20]. These models provide the *cdf* functions that are required for stochastic network design.

B. Related Work

The trade-off between various performance metrics in WSNs has been investigated previously. Applications involving multiple performance metrics are investigated in [3, 4, 14, 16, 21]. Specifically, the energy consumption and delay tradeoff problems are studied in [3, 4, 14, 21]. Trade-offs between connectivity, lifetime, and application-specific properties such as spatial density of sensing points are investigated in [5]. These studies share the same goal of exploring trade-offs in WSNs. However, only deterministic measures are considered and stochastic characteristics of the performance metrics are not captured.

Recent studies are focused on the probabilistic analysis of the delay [7, 8, 11], throughput [9, 22] and lifetime [13]. While they provide statistical information for the performance metrics of concern, interrelationship among different performance metrics are unexplored so far.

In our previous studies [17, 18, 20], the probability distribution of the end-to-end delay, the network lifetime, and the event detection delay are analyzed. While these studies lay the ground for the analysis in this paper, the tradeoffs among the performance metrics are left uninvestigated.

III. Problem Definition and Methodology

We consider a network characterized by three design parameters: the network density ρ, the locally generated traffic rate λ_{lc}, and the duty cycle ξ. The probabilistic performance metrics include the end-to-end delay from a node at the edge of the network to the sink, $ED = t_{\text{e2e}}(R)$, and the network lifetime, NL.

A. Quantile-based Design Measures

Knowledge of the probabilistic distribution of performance metrics provides extensive capabilities to network designers. To leverage these stochastic models, stochastic design measures are also necessary. To this end, we define two quantile-based design measures.

Consider a particular probabilistic performance metric $g(\boldsymbol{d})$, which can be either the end-to-end delay ED or the network lifetime NL, and is a function of a set of design parameters $\boldsymbol{d} = \{\rho, \lambda_{\rm lc}, \xi\}$. We define the following design measures:

Definition 1: **Network Dependability: The p-quantile** of a probabilistic performance metric $g(\boldsymbol{d})$, denoted by $g^{(p)}(\boldsymbol{d})$, is defined as the value of $g(\boldsymbol{d})$ achieved with at least a probability of p.

Definition 2: **Network Predictability: The (p_1, p_2)-quantile interval** of a probabilistic performance metric g, denoted by $g^{(p_1,p_2)}(\boldsymbol{d}), (p_1 \leq p_2)$, is defined as the difference between $g^{(p_1)}(\boldsymbol{d})$ and $g^{(p_2)}(\boldsymbol{d})$, i.e., $g^{(p_1,p_2)}(\boldsymbol{d}) = g^{(p_2)}(\boldsymbol{d}) - g^{(p_1)}(\boldsymbol{d}), (p_1 \leq p_2)$.

The p-quantile is the value of the performance metric with a probability guarantee, which can be denoted as the *dependability* of the network. The (p_1, p_2)-quantile interval is used to describe how the probabilistic performance metric is "concentrated", which can be denoted as the *predictability* of the network. For example, consider the $(0.1, 0.9)$-quantile interval of delay. A small interval suggests that for the majority of the packets (packets except the fastest 10% and the slowest 10%), the delay is concentrated in a small region between the 0.1-quantile and the 0.9-quantile. Thus, the delay performance of the network is easier to predict.

The p-quantile and (p_1, p_2)-quantile interval are directly obtained from the *cdf*s of corresponding performance metrics. Given a probabilistic metric $g(\boldsymbol{d})$, and its *cdf* $G_{g(\boldsymbol{d})}(g)$, the p-quantile and (p_1, p_2)-quantile interval are given by

$$g^{(p)}(\boldsymbol{d}) = G^{-1}_{g(\boldsymbol{d})}(p), \tag{3}$$

$$g^{(p_1,p_2)}(\boldsymbol{d}) = G^{-1}_{g(\boldsymbol{d})}(p_2) - G^{-1}_{g(\boldsymbol{d})}(p_1), \tag{4}$$

respectively, where $G_{g^{-1}(\boldsymbol{d})}(g)$ is the inverse *cdf*. Obtaining the closed-from inverse function for $G_{g(\boldsymbol{d})}(g)$ in practice may be infeasible. In our evaluations, a series of tuples $(g_1, p_1), (g_2, p_2), ...$ are obtained from the *cdf* $G_{g(\boldsymbol{d})}(g)$ based on models described in Section II-A. Then, $G^{-1}_{g(\boldsymbol{d})}(p)$ is obtained using spline interpolation.

B. Network Design Problem Formulation

Using the p-quantile and (p_1, p_2)-quantile interval measures, network design problem can be defined based on the type of objective function. More specifically, we consider two types of problems, where the objective function is a quantile measure or a quantile interval measure. Accordingly, the network design problem can be formulated as follows.

1) Quantile Objective Optimization: In this type of optimization problem, the objective function is the p-quantile of one of the probabilistic metrics, where p is an application-specific probability threshold:

$$\min_{\boldsymbol{d}} \; ED^{(p_{ed})}(\boldsymbol{d}) \quad \text{OR} \quad \max_{\boldsymbol{d}} \; NL^{(p_{nl})}(\boldsymbol{d}) \;, \tag{5}$$

given:

$$ED^{(p_{ed})} \leq ED_{\rm q}; \;\; NL^{(p_{nl})} \geq NL_{\rm q}, \tag{6}$$

$$ED^{(p_{ed1},p_{ed2})} \leq ED_{\rm v}; \;\; NL^{(p_{nl1},p_{nl2})} \leq NL_{\rm v}, \tag{7}$$

$$TP \geq TP_{\rm q}; \;\; d_{i1} \leq d_i \leq d_{i2}, (d_i \in \boldsymbol{d}), \tag{8}$$

where the probabilistic and deterministic constraints are given in (6)-(8), TP is the deterministic throughput, and (d_{i1}, d_{i2}) is the range for the design parameter d_i.

2) Quantile Interval Objective Optimization: In the second type of optimization problem, the objective function is the (p_1, p_2)-quantile interval of one of the probabilistic metric, where p_1 and p_2 are the application-specific probability thresholds:

$$\min_{\boldsymbol{d}} \; ED^{(p_{ed1},p_{ed2})}(\boldsymbol{d}) \;\; \text{OR} \;\; \min_{\boldsymbol{d}} \; NL^{(p_{nl1},p_{nl2})}(\boldsymbol{d}) \;, \tag{9}$$

given: constraints in (6) - (8).

C. Methodology

The solution to the above optimization problems is non-trivial because the performance measures are not convex functions and they cannot be converted to convex functions easily. For example, it can observed that the the 0.8-quantile of the network lifetime is non-convex with respect to the network density. The non-convexity of the solution space precludes the straight-forward use of standard optimization techniques, such as ILP. As our goal here is to study the stochastic performance tradeoffs in the design of WSNs, we use the following heuristic-based technique to solve the optimization problem and leave closed-form solutions to future work.

For an optimization problem defined by (5) or (9), we utilize a random initial search point methodology, where $N_{\rm search}$ local-optimum searches are conducted. In each of the multiple searches, the initial search point is determined by sequentially choosing random points within the parameter space, until one point falls within the feasible region. Starting from this point, a derivative-based local optimum search is conducted. Then, the global optimum is approximated by the best result in all the $N_{\rm search}$ optimum results found by each of the local searches. In the case when one or more of the local searches cannot converge due to non-convexity, these search procedures are terminated.

There are multiple benefits to utilizing this multiple-local-search technique. First, the technique does not require any form of prior knowledge about the topology and protocol. Second, the technique can be easily implemented taking advantage of multiple CPU cores or computers, since each of the local searches is totally independent of each other, thus can be executed in parallel. Finally, when $N_{\rm search}$ is large, the optimum found by this technique is asymptotically the global optimum, as the optimal solution eventually coincides with the local optimum related to one of the random initial points. It is easy to adjust the value of $N_{\rm search}$, such that a trade-off can be made between the accuracy of result and the computation time

efficiency. We also point out that the exact solution to the stochastic optimization problem is still an open issue and is out of the scope of this paper.

IV. Results

In this section, we first investigate the performance metrics; delay, throughput, and lifetime, as a function of the design parameters; traffic rate, network density, and duty cycle in Section IV-A. Then, in Section IV-B, we provide our main results for four key network design problems in WSNs through the optimization methodology described in Section III. The trade-offs between performance metrics are quantified.

The evaluations consider a network with a radius of $R = 30$m. The ranges of the design parameters in this study are as follows. The network density varies from $\rho = 0.004$ to 0.1 nodes/m^2; the traffic generation rate for each node ranges from $\lambda_{\rm lc} = 0.0004$ to 0.016 pkt/s; the duty cycle operation period is 10s, with a duty cycle ranging from 0.25 to 1. The time unit is chosen as 0.25s.

The data packet size is $l_p = 50$ bytes, whereas the beacon message and the CTS response message have the same size of $l_m = 22$ bytes. The beacon transmission timeout is $T_m = 10$ s. The channel related parameters (refer to [23] for detailed explanations) are listed as follows: the transmission power is set to -15 dBm for all the nodes. The threshold radius $r_{\rm th}$ is set to 10 m, within which all nodes only transmit packets to the sink. The SNR threshold is set to $\psi_{\rm th} = 10$ dB. Parameters for the channel are: $P_n = -105$ dBm, $P_L(D_0) = 52.1$ dB, $D_0 = 1$ m, $\eta = 3.3$, and $\sigma^s = 5.5$. This network setting results in multi hop paths of 3 - 10 hops.

A. Performance Metrics and Design Parameters

In the following, the analytical results of the relationship between performance metrics and design parameters are presented based on the models developed in [17, 20]. Our goal here is to investigate the characteristics of performance metrics within the parameter space and identify key trends. To the best of our knowledge, these performance metrics have not been investigated in a common network setting before.

In Figs. 2(a), and 2(b), the 0.9-delay is shown as a function of the traffic rate and network density, respectively[1]. As shown in Fig. 2(a), the 0.9-delay increases with traffic rate, since higher traffic rate causes higher queueing delay. Moreover, a lower network density causes the delay to increase because less nodes are active when each node starts to transmit. Thus, the waiting time is increased. For a very low network density ($\rho = 0.04$), it can be observed that the network cannot support a guaranteed 0.9-delay for traffic rates higher than 0.012 pkts/s. This is because less than 90% of the packets are delivered to the sink from the edge nodes. It is also shown that the 0.9-delay is generally a non-convex function of the traffic rate, motivating the need for the heuristic approach described in Section III.

[1]In the remaining part of this paper, when there is no ambiguity, we use p-delay and (p_1, p_2)-delay to represent the p-quantile and (p_1, p_2)-quantile interval of the end-to-end delay, and p-lifetime and (p_1, p_2)-lifetime for lifetime related measures.

Similarly, in Fig. 2(b), it is observed that when the network density, ρ, is less than 0.04 nodes/m^2, the 0.9-delay does not exist. This graph clearly shows the feasible region for a real-time WSNs and can be used as a guideline to determine network density.

For probabilistic network lifetime analysis, the relationship between the 0.9-lifetime and network density is shown in Fig. 2(c). The 0.9-lifetime has a peak when the density is around $0.15 - 0.3$ nodes/m^2, depending on the duty cycle. This is because when density is low, there is a higher chance that nodes are isolated from each other, and will spend more energy on continuously transmitting beacon messages. On the other hand, when the network density is higher, the total traffic forwarded to the sink is increased, thus the nodes close to the sink deplete their energy faster. The developed heuristic solution can be used to find the optimal density, as will be discussed in Section IV-B.

Next, we analyze the quantile values of delay. In Fig. 3(a), the p-delay for $p = [0.5, 0.7, 0.9]$ are shown as a function of the traffic rate. The three curves show the achievable end-to-end delay with these probabilities. In this evaluation, the network density is $\rho = 0.08$ node/m^2, and the duty cycle is $\xi = 0.2$. As a comparison, the average delay is also shown in the figure, which is calculated as

$$\bar{t}_{\rm e2e}(R) = \frac{\int_0^\infty t \cdot f_{\rm e2e}(R,t){\rm d}t}{\int_0^\infty f_{\rm e2e}(R,t){\rm d}t}, \tag{10}$$

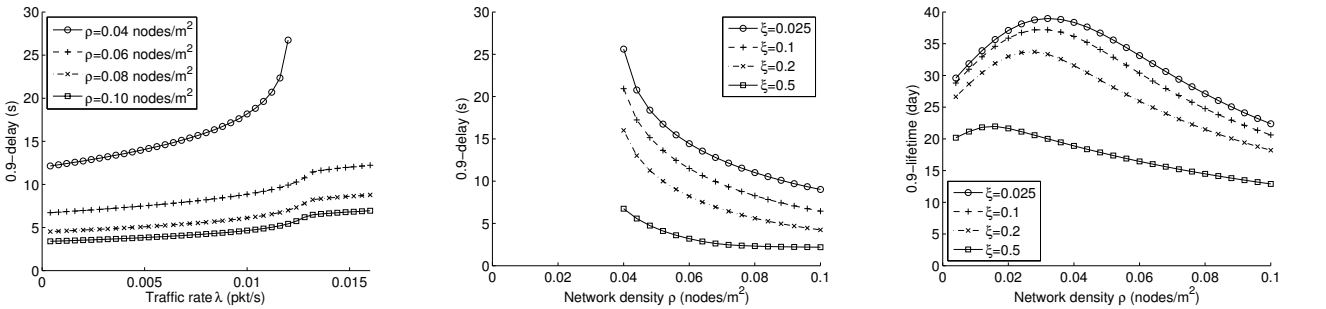
where $f_{\rm e2e}(R,t)$ is the *pdf* of the end-to-end delay from the edge of network to the sink.

In Fig. 3(a), the average delay has a similar trend w.r.t. the traffic rate as the 0.5- and 0.7-delay. However, as the traffic rate increases, the average delay grows slower than the quantile-based delay measures. As the traffic rate is increased, a larger portion of the packets is lost. As an example, for traffic rates higher than 0.006 pkt/s, more than 10% of the packets are lost. The average delay is calculated only for those packets that are eventually delivered. Therefore, the average delay does not contain the information of lost packets. Consequently, in high-rate and loss-tolerant applications, the average delay may lead to inaccurate design decisions.

B. Performance Trade-offs in Real-time WSNs

In this section, we first present the effectiveness of the heuristic optimization methodology. Then, we leverage this methodology to solve four main optimization problems to explore and quantify trade-offs in WSNs. We show the relationships between the optimal parameters and the performance requirements.

1) Evaluation of the Methodology: The solution to the probabilistic optimization problems is implemented using MATLAB. Each search procedure is conducted using the interior point method. In the case where local searches cannot converge, a limit on the number of iterations, MAX_ITER, is enforced. Accordingly, the search procedure is parametrized by the maximum iterations MAX_ITER and the number of searches N_{search}. For comparison, we utilize a discretized brute force search, where the objective function and constraint

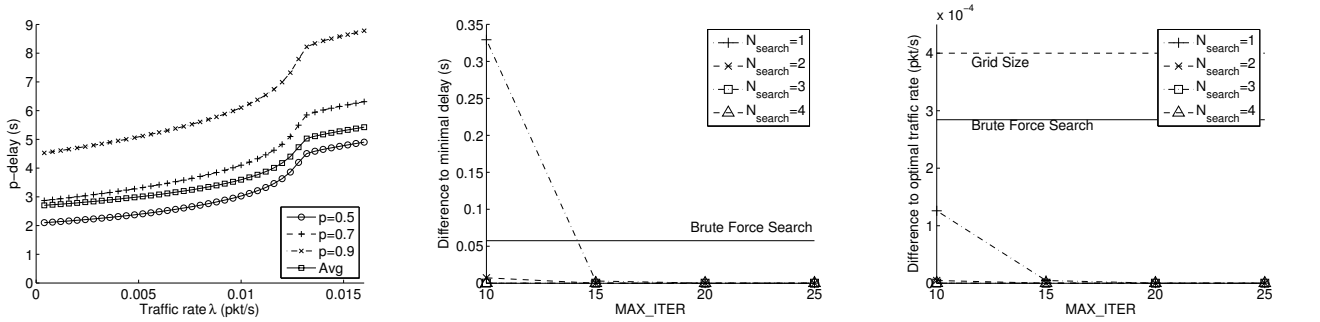


(a) 0.9-delay vs. traffic rate. Duty cycle is 0.2.

(b) 0.9-delay vs. network density. Traffic rate is 0.008 pkt/s.

(c) 0.9-lifetime vs. network density. Traffic rate is 0.008 pkt/s.

Fig. 2: Evaluation of performance metrics (delay, lifetime, and throughput) as a function of network design parameters (traffic rate, duty cycle, and network density).



(a) p-delay vs. traffic rate. Average delay is also shown.

(b) Minimum delay

(c) Optimal Traffic Rate

Fig. 3: (a) p-delay and (b)-(c) The difference between the benchmark solution and the solution in each setup. Solution of the brute force search is also shown.

functions are evaluated at grid points in the entire design parameter space. The ranges and increments of the three design parameters are selected as follows. The traffic rate, λ_{lc}, varies from 0.0004 pkt/s to 0.016 pkt/s with an increment of 0.0004 pkt/s; the duty cycle, ξ, 0.025-1 with 0.025 increments; and the network density, ρ, 0.004-0.1 nodes/m^2 with 0.004 nodes/m^2 increments.

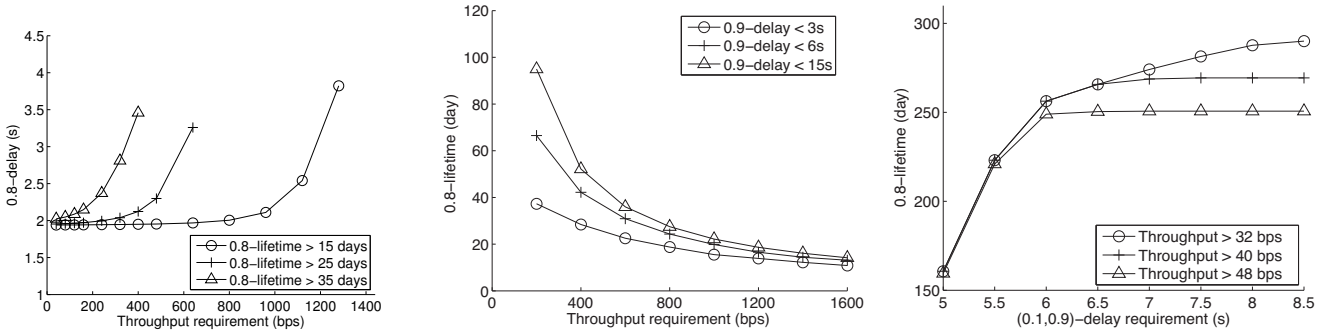
The evaluation of the objective and constraint function values at each point takes approximately $15-30$s and the total calculation time for all the $40,000$ points is approximately $7-14$ days. This delay prohibits much finer grid sizes and hence, limits the accuracy of the brute force search. In comparison, with the search methodology described in Section III, with 4 local searches and a maximum iteration of 25, the time needed is less than 2 hours.

Accuracy of the Multiple Local Search: We evaluate the multiple local search methodology for a delay minimization problem, where the same problem is solved with several choices of the number of searches, N_{search}, and maximum iteration allowed in each search, MAX_ITER. The results are compared to the discretized brute force search solution.

For evaluations, since a global optimum solution does not exist, we use the optimal solution found across all choices as the benchmark. The error of each choice (of N_{search} and MAX_ITER) is then represented as the difference in the resulting objective function or the optimum parameter from the benchmark. It remains an open problem to find the *exact* global optimal solution to the probabilistic optimization problems in this paper. The results, however, show that the majority of the resulting solutions converge to the same value.

In Fig. 3(b), the optimization error for the following optimization problem is shown: Minimize the 0.9-delay such that the throughput received at the sink is higher than 200 bps and the 0.8-lifetime is longer than 30 days. For each combination of N_{search} and MAX_ITER, 200 optimization procedures are conducted. Each procedure contains N_{search} local searches. Each of the 200 solutions is ordered based on its minimum delay value and the 0.9-quantile of the error is shown in Fig. 3(b). In other words, 90% of the solutions for each setup has an error equal to or smaller than the value shown in the y-axis.

The entire experiment contains $2,000$ local searches, out of which the best solution (error: 0, absolute value: 2.29 s) is obtained as the benchmark. For comparison, the brute force search result for all $40,000$ points in the parameter space is also shown (error: 0.0575 s, absolute value: 2.35 s). In all cases except when $N_{\text{search}} = 1$ and $MAX\_ITER \leq 15$, the multiple local search solution consistently yields better results than the brute force search. This is illustrated in Fig. 3(c), where the error in optimum traffic rate is shown. As can be seen, the error of the brute force search is less than the grid size (also shown), which suggests that the multiple iterative search finds the optimal solution in most cases.



(a) Minimum 0.8-delay (b) Maximum 0.8-lifetime (c) Maximum 0.8-lifetime

Fig. 4: Trade-offs between optimum probabilistic end-to-end delay, probabilistic network lifetime, and the throughput.

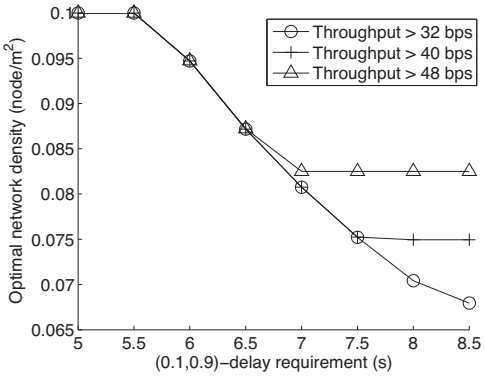


Fig. 5: Optimal network density for 4(c).

2) Probabilistic Performance Metric Tradeoffs: Next, the results of four optimization problems are presented based on the multiple local search approach discussed in Section III.

Quantile-based Delay Minimization: We first consider a stochastic version of a common problem in the design of WSNs: Minimize the quantile-based delay subject to throughput and lifetime requirements. More specifically, the minimum 0.8-delay is found with varying throughput and 0.8-lifetime requirements. In Fig. 4(a), the minimum achievable 0.8-delay is shown subject to a throughput requirement, which ranges from 40 bps to 1280 bps, and 0.8-lifetime requirement, which ranges from 15 days to 35 days. It can be observed that generally the minimum achievable 0.8-delay increases with higher throughput requirement and longer lifetime requirement. The well-known tradeoff between delay and throughput can be clearly quantified for WSNs. Moreover, independent of the lifetime and throughput requirements, delay is lower bounded, which is dominated by the network topology (i.e., density and duty cycle).

Quantile-based Lifetime Maximization: Maximizing network lifetime is essential for the proliferation of WSNs. Moreover, providing stochastic lifetime guarantees is crucial since each deployment may result in a different realization. To this end, our goal is to maximize the 0.8-lifetime subject to throughput and 0.9-delay requirements. The results for the achievable maximum 0.8-lifetime is shown in Fig. 4(b). It can be observed that an increase in the throughput requirement decreases network lifetime with diminishing effects. Moreover, for relatively high-throughput applications ($TP > 1.2kbps$), a slight relaxation of the lifetime requirement can significantly improve the delay performance. For example, for a high throughput requirement of 1.6kbps, an 8% (24%) relaxation of the lifetime requirement improves the 0.9-delay by 60% from 15s to 6s (by 80% to 3s). On the other hand, for a low throughput requirement of 200 bps, the same improvement require a relaxation of 29% (60%) in the lifetime requirement.

Quantile Interval Constraints: In most real-time applications, predictable delay performance is more important than minimizing delay as the task model can be approximated by a deterministic one. Consequently, it is important to constrain the quantile interval measure of delay in these cases. Thus, we consider a network lifetime maximization problem, where the 0.8-lifetime is maximized subject to a throughput requirement, which ranges from 32 to 48 bps, the $(0.1, 0.9)$-delay requirement, which ranges from 6s to 9s and a 0.9-delay requirement of 15s.

The maximum achievable 0.8-lifetime is shown in Fig. 4(c). The 0.8-lifetime decreases when a lower (tighter) $(0.1, 0.9)$-delay or a higher throughput is required. On the other hand, to prolong lifetime, either delay predictability or throughput has to be sacrificed. An important finding from the figure is the 0.8-lifetime is dominantly determined by the $(0.1, 0.9)$-delay requirement when the quantile interval is less than 6s. When the quantile interval is higher than 7s, the 0.8-lifetime is dominantly determined by the throughput requirement.

The optimal design parameters for the third scenario are also examined. In Fig. 5, the optimal network density is shown as a function of the $(0.1, 0.9)$-delay requirement, and the throughput requirement, corresponding to the maximum 0.8-lifetime. The right part of the figure ($(0.1, 0.9)$-delay > 7 s) shows that when the $(0.1, 0.9)$-delay requirement is relatively high (relaxed), the optimal network density is mainly determined by the throughput requirement and is independent of the $(0.1, 0.9)$-delay requirement. As shown in Fig. 4(c), in this region, the lifetime cannot be significantly increased. In the middle of the Fig. 5 (5.5 s $<$ $(0.1, 0.9)$-delay $<$ 7 s), the optimum network density increases when the $(0.1, 0.9)$-delay requirement is tightened, regardless of the throughput requirement. Therefore, the $(0.1, 0.9)$-delay requirement is the dominant requirement is this region. Finally, in the left part of the figure,

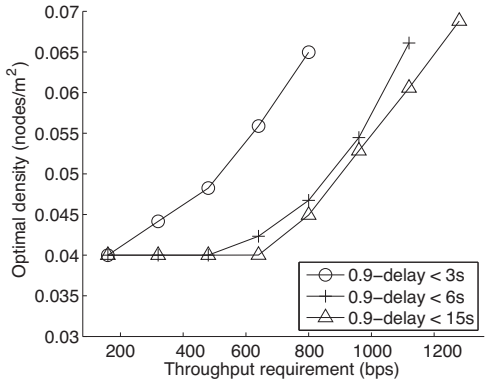


Fig. 6: Minimum network density as a function of throughput requirement and 0.9-delay requirement.

the density cannot be further increased, because 0.1 node/m^2 is the highest density in our parameter space, representing a upper limit on the deployment cost of the network. Therefore, the design parameter range is the dominant factor and the maximum achievable (0.8)-lifetime is significantly reduced in this region (Fig. 4(c)).

Network Cost Minimization: In the last scenario, we consider a network density minimization, i.e., network cost minimization, problem subject to a 0.9-delay requirement, which ranges from 3s to 15s, a throughput requirement, which ranges from 160bps to 1280bps, a 0.8-lifetime requirement of 15 days. The optimal density as a function of throughput requirement is shown in Fig. 6. It can be observed that a lower (relaxed) throughput requirement (160bps) results in the lowest density. An increase in the throughput requirement to 640bps, leads to an increase in the optimum traffic rate without affecting the optimum density. A further increase in the throughput requirement also requires a higher density. Accordingly, when the throughput requirement is higher than 640 bps, the optimal density is dominantly determined by the throughput requirement (the solution resides on the throughput requirement boundary), but when the throughput requirement is lower, the optimal density is dominantly determined by the end-to-end delay requirement (the optimal solutions reside on the 0.9-delay requirement boundary).

The impact of delay requirement on network design can be observed in Fig. 6. As the 0.9-delay requirement is relaxed from 3s to 15s, a lower network density can be allowed while still guaranteeing the throughput requirement. For example, for a throughput requirement of 640 bps, when 0.9-delay requirement is decreased from 15s to 6s, the optimum density is increased by only 6%. On the other hand, a further improvement of the 0.9-delay requirement to 3s requires a 40% higher network density. These results highlight important trade-offs for efficient network design.

V. Conclusions and Future Work

In this paper, the trade-offs of probabilistic performances metrics for real-time WSNs are explored and quantified. The trade-offs are investigated by formulating probabilistic optimization problems and the solutions are found using a heuristic-based technique. Two probabilistic performance measures are developed to characterize the dependability and predictability of the performances.

References

[1] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921–960, Mar 2007.

[2] C. Bin, C. Linlin, Y. Minghua, G. Taolin, and T. Chengping, "Da-mac: A duty-cycled, directional adaptive mac protocol for airborne mobile sensor network," in *Digital Manufacturing and Automation (ICDMA), 2013 Fourth International Conference on*, June 2013, pp. 389–392.

[3] N. Boughanmi and Y. Song, "A new routing metric for satisfying both energy and delay constraints in wireless sensor networks," *Journal of Signal Processing Systems*, vol. 51, no. 2, pp. 137–143, 2008.

[4] I. Demirkol and C. Ersoy, "Energy and delay optimized contention for wireless sensor networks," *Computer Networks*, vol. 53, no. 12, pp. 2106–2119, Aug 2009.

[5] K. Ferentinos and T. Tsiligiridis, "Adaptive design optimization of wireless sensor networks using genetic algorithms," *Computer Networks*, vol. 51, no. 4, pp. 1031–1051, Mar 2007.

[6] J. Kim, X. Lin, and N. B. Shroff, "Optimal anycast technique for delay-sensitive energy-constrained asynchronous sensor networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 2, pp. 484–497, 2011.

[7] J. Lehoczky, "Real-time queueing network theory," in *Proc. of IEEE RTSS 1997*, San Francisco, CA, Dec 1997, pp. 58–67.

[8] H. Li, P. Shenoy, and K. Ramamritham, "Scheduling messages with deadlines in multi-hop real-time sensor networks," in *Proc. of IEEE RTAS 2005*, San Francisco, CA, Mar 2005, pp. 415–425.

[9] R. Lubben, M. Fidler, and J. Liebeherr, "Stochastic bandwidth estimation in networks with random service," *IEEE/ACM Trans. Networking*, vol. 22, no. 2, pp. 484–497, Apr. 2014.

[10] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: an Algorithmic Approach*. Dover Publications Inc., 1981.

[11] M. Neuts, J. Guo, M. Zukerman, and H. L. Vu, "The waiting time distribution for a TDMA model with a finite buffer and state-dependent service," *IEEE Trans. on Communications*, vol. 53, no. 9, pp. 1522–1533, Sep 2005.

[12] Y. Noh, U. Lee, P. Wang, B. S. C. Choi, and M. Gerla, "VAPR: void-aware pressure routing for underwater sensor networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 5, pp. 895–908, May 2013.

[13] M. Noori and M. Ardakani, "Lifetime analysis of random event-driven clustered wireless sensor networks," *IEEE Trans. Mob. Comput.*, vol. 10, no. 10, pp. 1448–1458, Dec 2011.

[14] L. Shi and A. Fapojuwo, "TDMA scheduling with optimized energy efficiency and minimum delay in clustered wireless sensor networks," *IEEE Trans. on Mobile Computing*, vol. 9, no. 7, pp. 927–940, July 2010.

[15] M. C. Vuran and I. F. Akyildiz, "XLP: A cross layer protocol for efficient communication in wireless sensor networks," *IEEE Trans. on Mobile Computing*, vol. 9, no. 11, pp. 1578–1591, Nov 2010.

[16] D. Wang, B. Xie, and D. Agrawal, "Coverage and lifetime optimization of wireless sensor networks with gaussian distribution," *IEEE Trans. on Mobile Computing*, vol. 7, no. 12, pp. 1444–1458, Dec 2008.

[17] Y. Wang, M. C. Vuran, and S. Goddard, "Stochastic analysis of energy consumption in wireless sensor networks," in *Proc. of IEEE SECON 2010*, Boston, MA, Jun 2010.

[18] ——, "Analysis of event detection delay in wireless sensor networks," in *Proc. of IEEE INFOCOM 2011*, Shanghai, China, Apr 2011.

[19] ——, "Stochastic analysis of energy consumption in wireless sensor networks," University of Nebraska-Lincoln, http://cse.unl.edu/~yunbow/pdf/energy_dist_tr.pdf, Tech. Rep. TR-UNL-CSE-2011-0016, Jul 2011.

[20] ——, "Cross-layer analysis of the end-to-end delay distribution in wireless sensor networks," *IEEE Trans. on Networking*, vol. 20, no. 1, pp. 305–318, Feb 2012, (an earlier version appeared in IEEE RTSS 2009, Dec. 2009).

[21] Y. Yun and Y. Xia, "Maximizing the lifetime of wireless sensor networks with mobile sink in delay-tolerant applications," *Mobile Computing, IEEE Transactions on*, vol. 9, no. 9, pp. 1308–1318, 2010.

[22] K. Zheng, F. Liu, L. Lei, and Y. Jiang, "Stochastic performance analysis of a wireless finite-state markov channel," *IEEE Trans. Wireless Communications*, vol. 12, no. 2, pp. 782–793, Feb. 2013.

[23] M. Zúniga and B. Krishnamachari, "An analysis of unreliability and asymmetry in low-power wireless links," *ACM Trans. on Sensor Networks*, vol. 3, no. 2, June 2007.